Cauchy Noise Removal by Nonconvex ADMM with Convergence Gaurantees

Mei, Jin-Jin; Dong, Yiqiu; Huang, Ting-Zhu ; Yin, Wotao

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Cauchy Noise Removal by Nonconvex ADMM with Convergence Guarantees *

Jin-Jin Mei† Yiqiu Dong‡ Ting-Zhu Huang§ Wotao Yin¶

Abstract

Image restoration is one of the most important and essential issues in image processing. Cauchy noise in engineering application has the non-Gaussian and impulsive property. In order to preserve edges and details of images, the total variation (TV) based variational model has been studied for restoring images degraded by blur and Cauchy noise. Due to the nonconvexity and nonsmoothness, there exist computational and theoretical challenges. In this paper, adapting recent results, we develop an alternating direction method of multiplier (ADMM) in spite of the challenges. The convergence to a stationary point is guaranteed theoretically under certain conditions. Experimental results demonstrate that the proposed method is competitive with other methods in terms of visual and quantitative measures. Especially, by comparing to the PSNR values, our method can improve about 0.5dB on average.

Key words: nonconvex variational model, image restoration, total variation, alternating direction method of multiplier, Kurdyka-Łojasiewicz

1 Introduction

In many practical applications, the images inevitably mix some natural noises with the non-Gaussian property, such as impulse noise, Poisson noise, multiplicative noise and Cauchy noise. At the same time, the images often be blurred by the point spread function (PSF) during the acquisition and transmission. Therefore, the image restoration problem is one of the most important and essential image processing tasks in applied mathematics. The researchers have proposed many algorithmic methods to restore the blurry and noisy images; see [12, 16, 17, 27, 35, 36, 41, 54] and references therein. In this paper, we focus on recovering the blurred images corrupted by Cauchy noise. This kind of noise usually arises in echo of radar, low-frequency atmospheric noises and underwater acoustic signals [26, 31, 40]. According to [44, 45], it follows Cauchy distribution and possesses impulsive character.

*This research was supported by 973 Program (2013CB329404), NSFC (61370147, 61402082, 11401081), the Fundamental Research Funds for the Central Universities (ZYGX2013Z005, ZYGX2013J106).

†School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, People's Republic of China, and Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark (meijinjin666@126.com).

‡Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark (yido@dtu.dk). The work was supported by Advanced Grant 291405 from the European Research Council.

§School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, People's Republic of China (tingzhuhuang@126.com).

¶Department of Mathematics, University of California-Los Angeles, Los Angeles, CA 90025, USA (wotaoyin@math.ucla.edu). The work was supported by NSF Grant ECCS-1462398 and ONR Grant N000141410683.

Without loss of generality, we assume that the original gray-scale image u is defined on a connected bounded space $\Omega \subset \mathbb{R}^2$ with compacted Lipschitz boundary. The observed image with blur and Cauchy noise is given as follows:

$$f = Ku + \eta, \tag{1}$$

where $f \in L^2(\Omega)$ denotes the observed image, $K \in \mathcal{L}(L^1(\Omega), L^2(\Omega))$ represents a known linear and continuous blurring (or convolution) operator and $\eta \in L^2(\Omega)$ denotes Cauchy noise. Here we aim at recovering u from the observed image f, which is an ill-posed inverse problem.

In recent years, more attentions have been paid on Cauchy noise, and several related denoising methods have been proposed. In [13], the authors applied a recursive algorithm based on Markov random field to reconstruct images and retain the sharp edges. In 2005, Achim and Kuruoğlu utilized a bivariate maximum a posteriori estimator (BMAP) to propose a new statistical model in the complex wavelet domain for removing Cauchy noise [1]. In [34], Loza et al. proposed a statistical approach based on non-Gaussian distribution in the wavelet domain for tackling the image fusion problems. Their method achieved an obvious improvement with respect to the fusion quality and the noise reduction. In [46], Wan et al. developed a novel image segmentation method for RGB images corrupted by Cauchy noise. They combined the statistical methods with the denoising techniques to obtain the satisfactory performance. In 2015, Sciacchitano et al. proposed a new total variation (TV) based variational method for recovering Cauchy noise corrupted images [42]. Due to the use of the TV regularization, the fine features and the important edges of images can be preserved effectively [21]. But, since the data-fitting term based on Cauchy distribution is nonconvex, the solutions of the corresponding optimisation problem strongly depend on initializations and numerical schemes. To overcome this difficulty, based on the median filter's result by adding a quadratic penalty term, in [42] the new convex variational model is given as

$$\min_{u \in BV(\Omega)} \int_\Omega |Du| + \frac{\lambda}{2}\left(\int_\Omega \log\left(\gamma^2 + (u-f)^2\right) dx + \alpha \|u - \tilde{u}\|_2^2\right), \tag{2}$$

where $\gamma > 0$ is the scale parameter of Cauchy distribution, and $BV(\Omega)$ is the space of functions of bounded variation, i.e. $u \in BV(\Omega)$ iff $u \in L^1(\Omega)$ and its total variation (TV)

$$\int_\Omega |Du| = \sup\left\{ \int_\Omega u \, \mathrm{div}\vec{v} \, dx : \vec{v} \in (C_0^\infty(\Omega))^2, \|\vec{v}\|_\infty \le 1 \right\}$$

is finite. Here, $(C_0^\infty(\Omega))^2$ is the space of vector-valued functions with compact support in Ω. The space $BV(\Omega)$ endowed with the norm $\|u\|_{BV(\Omega)} = \|u\|_{L^1(\Omega)} + \int_\Omega |Du|$ is a Banach space; see, e.g., [21]. In (2), λ denotes the positive regularization parameter, which achieves the trade-off between the smoothness from TV regularization and a good fit to f and $\tilde{u}$, $\tilde{u}$ is the result obtained by the median filter, and α is a positive penalty parameter. Note that if $8\alpha\gamma^2 \ge 1$, the objective functional in (2) is strictly convex and there exists a uniqueness solution. Then, the primal-dual algorithm [11] is utilized to solve the convex minimization problem (2). The last term in (2) is in fact pushing the solution close to the median filter results, but the median filter does not always provide satisfactory results on removing Cauchy noise. Hence, in this paper we turn our focus back to the nonconvex model to discover its performance and properties.

Recently, researchers have discovered some convergence properties of the optimization algorithms for solving nonconvex minimization problems [23, 47, 48, 53]. In particular, the paper [48] introduced the global convergence of the alternating direction method of multipliers (ADMM) for nonconvex nonsmooth optimization with linear constraints. To take advantages of the recent results, in this paper we develop the ADMM algorithm to solve the nonconvex variational model

directly for denoising and deblurring simultaneously, i.e.,

$$\min_{u\in BV(\Omega)} \int_\Omega |Du| + \frac{\lambda}{2}\int_\Omega \log\left(\gamma^2 + (Ku-f)^2\right) dx. \tag{3}$$

Then, we prove that the ADMM algorithm starting from any initialization is globally convergent to a stationary point under certain conditions. Furthermore, comparisons with the state-of-the-art method proposed in [42] show the effectiveness of our proposed method in terms of the image restoration quality and noise reduction.

The outline of the paper is summarized as follows. In the next section, we analyse some fundamental properties of Gaussian distribution, Laplace distribution and Cauchy distribution. In Section 3, we illustrate the nonconvex variational model for denoising and deblurring, and prove the existence and uniqueness of the solution. In Section 4, we develop the ADMM algorithm for the proposed nonconvex model and present the convergence results. In Section 5, we demonstrate the performance of the ADMM algorithm on the nonconvex model (3) by comparing with other existing algorithms. Finally, we conclude the paper with some remarks in Section 6.

2 Statistical Properties for Cauchy Distribution

Cauchy distribution is a special case of the α-stable distribution with $\alpha = 1$, and is important as a canonical example of the "pathological" case [3, 15, 29]. It is closed under linear fractional transformations with real coefficients [30]. However, different from the most α-stable distributions, it possesses a probability density function that can be expressed analytically [19, 28]:

$$P_C(x) = \frac{\gamma}{\pi\left((x-\mu)^2+\gamma^2\right)},$$

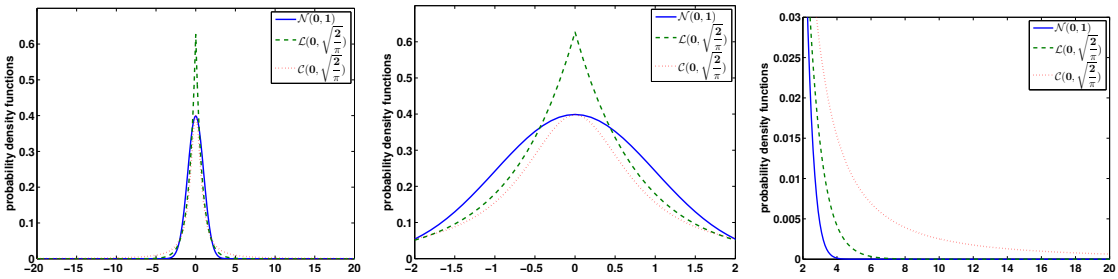
where μ is the parameter to specify the location of the peak, $\gamma > 0$ is the scale parameter that decides the half-width at half-maximum. Here, we use $C(\mu, \gamma)$ to denote Cauchy distribution. Its mode and median are both μ, while the mean, variance and higher moments are undefined. In addition, Cauchy distribution is infinitely divisible, i.e., for every positive integer n, there exists n independent identically distribution (i.i.d.) random variables $X_{n1}, X_{n2}, \ldots X_{nn}$ satisfied that $X_{n1} + X_{n2} + \cdots + X_{nn}$ follows Cauchy distribution. Due to the property of the infinite divisibility, random variables following Cauchy distribution obey the generalized central limit theorem [37].

Cauchy distribution is closely related to some other probability distributions. Cauchy distribution belongs to the heavy-tailed distribution, where the tail heaviness is determined by the scale parameter γ. In particularly, if X and Y are two independent Gaussian distributed random variables with mean 0 and variance 1, then the ratio X/Y follows with the standard Cauchy distribution, i.e., $C(0, 1)$ [6, 38]. Therefore, we determine to apply this property to create the noisy images corrupted by Cauchy noise in the following Section 5.

Further to show the statistical properties of the Cauchy distribution, we compare it with two most common used probability distributions: Gaussian distribution ($\mathcal{N}(\mu, \sigma^2)$ with mean μ and variance σ^2) and Laplace distribution ($\mathcal{L}(\mu, b)$ with mean μ and variance $2b^2$).

Since both Gaussian distribution and Cauchy distribution belong to α-stable distribution with $\alpha = 2$ and $\alpha = 1$, respectively, they both are bell-shaped. Moreover, we can easily obtain the following relation between them at $x = 0$.

Proposition 2.1. *Let X be a random variable. Assume that $X \sim \mathcal{N}(0, 1)$. Then X has the same probability as $C(0, \sqrt{\frac{2}{\pi}})$ at $x = 0$.*



(a) (b) (c)

Figure 1. Comparison for probability density functions of $\mathcal{N}(0,1)$, $\mathcal{L}(0,\sqrt{\frac{2}{\pi}})$ and $C(0,\sqrt{\frac{2}{\pi}})$. (a) The plots of three distributions, (b) the zoomed-in portion of the curves around the peaks, (c) the zoomed-in portion of the curves around the tails.

In addition, both Laplace distribution and Cauchy distribution belong to the heavy-tailed distributions. We demonstrate the mutual relation among them around the tails of distribution curves in the following proposition.

Proposition 2.2. *Let X be a random variable. Assume that $\mu = 0$, we have the following relations:*

1. *If $x = \sigma = b = \gamma$, the ratio of P_G, P_L and P_C is $1 : \sqrt{\frac{\pi}{2e}} : \sqrt{\frac{e}{2\pi}}$.*

2. *If $x = 3\sigma = 3b = 3\gamma$, the ratio of P_G, P_L and P_C is $1 : \sqrt{\frac{\pi}{2}}e^{\frac{3}{2}} : \sqrt{\frac{1}{50\pi}}e^{\frac{9}{2}}$,*

where P_G and P_L denote the probability density functions for $\mathcal{N}(0,\sigma^2)$ and $\mathcal{L}(0,b)$.

Based on Proposition 2.2, for these three specific distributions, we can see that a random variable with Gaussian distribution has largest probability to fall at $x = \sigma = b = \gamma$, which corresponds to small perturbation. But at $x = 3\sigma = 3b = 3\gamma$, which corresponds to high perturbation, the probability from Laplace distribution is 5 times more than that from Gaussian distribution, and the probability from Cauchy distribution is even 7 times more.

In Figure 1, we show the curves of three probability density functions of Gaussian distribution, Laplace distribution and Cauchy distribution. From Figure 1(a), we find that these three distributions have different behaviours at the peaks and tails, see the details in the zoomed-in portions. Figure 1(b) shows the zoomed-in portion around the peaks of three distributions. Gaussian distribution has the same peak compared with Cauchy distribution while the curve of Gaussian distribution is slightly higher on both sides of the peak. Figure 1(c) shows the zoomed-in portion around the tails of three distributions. The tail of Laplace distribution is closer to Cauchy distribution compared with Gaussian distribution. However, there still exists a gap between the curves of Laplace distribution and Cauchy distribution. Therefore, Cauchy distribution can not be simply replaced with Gaussian distribution or Laplace distribution in image restoration problems.

3 Nonconvex Variational Model

In this paper, we focus on the deblurring and denoising case. In [42], a variational model for the denoising case was proposed. For self-containedness, similarly we deduce the nonconvex variational model for the deblurring and denoising case based on the maximum a posteriori (MAP) estimator and the Bayes's rule. Subsequently, we prove its existence and uniqueness.

3.1 Nonconvex variational model via MAP estimator

We consider $f(x)$ and $u(x)$ for each $x \in \Omega$ as random variables. The maximum a posteriori (MAP) estimate of u is the most likely value of u given f, i.e., $u^* = \arg\max_u P(u|f)$. Based on Bayes's rule and the independence of $u(x)$ and $f(x)$ for all $x \in \Omega$, we obtain

$$\begin{aligned} \max_u P(u|f) &\Longleftrightarrow \max_u \frac{P(f|u)P(u)}{P(f)} \\ &\Longleftrightarrow \min_u -\log P(f|u) - \log P(u) \\ &\Longleftrightarrow \min_u -\int_\Omega \log P(f(x)|u(x))\, dx - \log P(u), \end{aligned} \tag{4}$$

where the term $\log P(f(x)|u(x))$ describes the degradation process that produced f from u based on (1), and the term $\log P(u)$ is called as the prior on u. Since η follows $C(0,\gamma)$ for each $x \in \Omega$, we have

$$P(f(x)|u(x)) = \frac{\gamma}{\pi\left(((Ku)(x) - f(x))^2 + \gamma^2\right)}.$$

In addition, we use the prior $P(u) = \exp(-\mu \int_\Omega |Du|)$. Then, we obtain the variational model for the deblurring and denoising case

$$\min_{u \in BV(\Omega)} \int_\Omega |Du| + \frac{\lambda}{2} \int_\Omega \log\left(\gamma^2 + (Ku - f)^2\right)\, dx, \tag{5}$$

where $\lambda = \frac{2}{\mu} > 0$ is the regularization parameter. Although the TV regularization is convex, due to the logarithm in the data-fitting term, $\int_\Omega \log\left(\gamma^2 + (Ku - f)^2\right)\, dx$ is nonconvex. Therefore, the solutions of (5) strongly depend on the initializations and numerical schemes.

3.2 Existence and uniqueness of a solution to the model (5)

According to the properties of the total variation, we prove that there exists at least one solution for the nonconvex variational problem in the BV space.

Theorem 3.1. *Assume that Ω is a connected bounded set with compacted Lipschitz boundary and $f \in L^2(\Omega)$. Suppose that $K \in \mathcal{L}(L^1(\Omega), L^2(\Omega))$ is nonnegative and linear with $K\mathbf{1} \neq 0$. Then the model (5) has at least one solution $u^* \in BV(\Omega)$.*

Proof. Set $E(u) = \int_\Omega |Du| + \frac{\lambda}{2} \int_\Omega \log\left(\gamma^2 + (Ku - f)^2\right)\, dx$. Obviously, $E(u)$ is bounded from below. Considering a minimizing sequence $\{u^k\}$, we know that $E(u^k)$ is bounded, so we have that both $\left\{\int_\Omega |Du^k|\right\}$ and $\int_\Omega \log\left(\gamma^2 + (Ku^k - f)^2\right)\, dx$ are bounded.

Since the logarithmic function is strictly increasing and $\gamma > 0$, we can easily obtain that $\{Ku^k\}$ is bounded in $L^2(\Omega)$ and in $L^1(\Omega)$.

Based on $\left\{\int_\Omega |Du^k|\right\}$ bounded, by using the Poincaré inequality [2], we have

$$\|u^k - m_\Omega(u^k)\|_2 \leq C \int_\Omega |D(u^k - m_\Omega(u^k))| = C \int_\Omega |Du^k|, \tag{6}$$

where $m_\Omega(u^k) = \frac{1}{|\Omega|} \int_\Omega u^k dx$, C is a positive constant and $|\Omega|$ represents the measure of Ω. As Ω is bounded, then $\|u^k - m_\Omega(u^k)\|_2$ and $\|u^k - m_\Omega(u^k)\|_1$ are bounded for each k. Because $K \in \mathcal{L}(L^1(\Omega), L^2(\Omega))$ is continuous, we have that $\{K(u^k - m_\Omega(u^k))\}$ is bounded in $L^2(\Omega)$ and $L^1(\Omega)$. Thus, we have that

$$|m_\Omega(u^k)| \cdot \|K\mathbf{1}\|_1 \leq \|K(u^k - m_\Omega(u^k))\|_1 + \|Ku^k\|_1.$$

Due to $K\mathbf{1} \neq 0$, we get that $m_\Omega(u^k)$ is uniformly bounded. Combining with (6), the sequence $\{u^k\}$ is bounded in $L^2(\Omega)$ and in $L^1(\Omega)$. Recalling that $\left\{\int_\Omega |Du^k|\right\}$ is bounded, we get the boundedness of $\{u^k\}$ in $BV(\Omega)$.

Therefore, there exists a subsequence $\{u^{n_k}\}$ that converges strongly in $L^1(\Omega)$ to some $u^* \in BV(\Omega)$ as $k \to \infty$, while $\{Du^{n_k}\}$ converges weakly as a measure to Du^*. Since K is linear and continuous, $\{Ku^{n_k}\}$ converges strongly to Ku^* in $L^2(\Omega)$. According to the lower semicontinuity of the total variation and Fatou's lemma, we obtain that u^* is a solution of the model (5). □

Although the objective function in (5) is nonconvex, we are still able to obtain a result on the uniqueness of the solution under some assumptions.

Theorem 3.2. *Assume that $f \in L^2(\Omega)$ and K is injective, then the model* (5) *has a unique solution u^* in $\Omega_U := \{u \in BV(\Omega) : f(x) - \gamma < (Ku)(x) < f(x) + \gamma \text{ for all } x \in \Omega\}$.*

Proof. For each fixed $x \in \Omega$, we define a function $g : \mathbb{R} \to \mathbb{R}$ as

$$g(t) = \log\left(\gamma^2 + (t - f(x))^2\right).$$

Based on the second order derivative of g:

$$g''(t) = \frac{2\left(\gamma^2 - (t - f(x))^2\right)}{\left(\gamma^2 + (t - f(x))^2\right)^2},$$

we get that g is strictly convex, if $f(x) - \gamma < t < f(x) + \gamma$. Since K is injective, we have that if $f(x) - \gamma < (Ku)(x) < f(x) + \gamma$, $g((Ku)(x))$ is strictly convex. Combining with the fact that the TV is convex and K is a linear operator, the objective function of the model (5) is strictly convex in Ω_U. Hence, there exists a unique solution for the model (5) in Ω_U. □

Note that Cauchy noise is so impulsive even with small γ that many points in f are heavily corrupted. Then, the images in Ω_U should still have some impulsive noise left. If we also take the smoothing property of the blurring operator K into account, the unique solution in Ω_U will not be a satisfactory restoration result. In Section 5.1, we will show it numerically.

4 Proposed ADMM Algorithm

Due to the nonconvexity, the solutions of the variational model (5) strongly depend on the initializations and numerical algorithms. Taking advantage of the recent results in [48], in this section we apply the ADMM algorithm to solve the minimization problem in (5) for restoring the degraded images with blurring and Cauchy noise. Then, we prove that the proposed ADMM algorithm is globally convergent to a stationary point under certain conditions.

4.1 The ADMM algorithm for nonconvex and nonsmooth problem

Before applying the ADMM algorithm to solve the nonconvex model (5), we briefly review this algorithm and its convergence result under the nonconvex and nonsmooth case in [48].

Let $\mathbf{x} = [x_1^\top, \cdots, x_s^\top]^\top \in \mathbb{R}^N$ and $\mathbf{A} = [A_1, \cdots, A_s] \in \mathbb{R}^{M\times N}$ where $x_i \in \mathbb{R}^{n_i}$, $A_i \in \mathbb{R}^{M\times n_i}$, $\sum_{i=1}^s n_i = N$. We consider the minimization problem formulated as,

$$\min_{\mathbf{x},y} \mathcal{F}(\mathbf{x}) + \mathcal{G}(y) \quad \text{s.t. } \mathbf{A}\mathbf{x} + By = 0 \tag{7}$$

where $\mathcal{F}(\mathbf{x})$ is a continuous, proper, possibly nonsmooth and nonconvex function, $\mathcal{G}(y)$ is a proper, differentiable and possibly nonconvex function, and $y \in \mathbb{R}^L$ is a variable with the corresponding coefficient $B \in \mathbb{R}^{M\times L}$. By introducing a Lagrangian multiplier $w \in \mathbb{R}^M$ for the linear constraint $\mathbf{A}\mathbf{x} + By = 0$, we have the augmented Lagrangian function

$$\mathcal{L}_\beta(\mathbf{x}, y; w) = \mathcal{F}(\mathbf{x}) + \mathcal{G}(y) + w^\top(\mathbf{A}\mathbf{x} + By) + \frac{\beta}{2}\|\mathbf{A}\mathbf{x} + By\|_2^2,$$

where β is a positive penalty parameter.

According to the framework of the ADMM algorithm [7, 20], the solution $(\mathbf{x}^{k+1}, y^{k+1})$ of the nonconvex optimization can be achieved iteratively:

$$\begin{cases} x_1^{k+1} = \arg\min\limits_{x_1} \mathcal{L}_\beta(x_1, x_2^k, \cdots, x_s^k, y^k; w^k) \\ \quad\vdots \quad = \quad\vdots \qquad\vdots \\ x_s^{k+1} = \arg\min\limits_{x_s} \mathcal{L}_\beta(x_1^{k+1}, x_2^{k+1}, \cdots, x_s, y^k; w^k) \\ y^{k+1} = \arg\min\limits_{y} \mathcal{L}_\beta(\mathbf{x}^{k+1}, y; w^k) \\ w^{k+1} = w^k + \beta(\mathbf{A}\mathbf{x}^{k+1} + By^{k+1}). \end{cases} \tag{8}$$

The convergence result of the ADMM algorithm under the nonconvex and nonsmooth case is as follows [48].

Theorem 4.1. *Let $\mathcal{D} = \left\{(\mathbf{x}, y) \in \mathbb{R}^{N+L} : \mathbf{A}\mathbf{x} + By = 0\right\}$ be the nonempty feasible set and $\mathcal{F}(\mathbf{x})+\mathcal{G}(y)$ is coercive over $\mathcal{D}$, i.e., $\mathcal{F}(\mathbf{x}) + \mathcal{G}(y) \to \infty$ for any $(\mathbf{x}, y) \in \mathcal{D}$ and $\|(\mathbf{x}, y)\| \to \infty$. Suppose that $\mathbf{A}$, B have full column rank [1] and $Im(\mathbf{A}) \subset Im(B)$. Assume that $\mathcal{F}(\mathbf{x})$ is either restricted prox-regular [2] or piecewise linear, and $\mathcal{G}(y)$ is Lipschitz differentiable with the constant $L_{\nabla\mathcal{G}} > 0$. Then, the ADMM algorithm is subsequentially convergent for any β larger than a certain constant β_0. If $\mathcal{L}_\beta$ satisfies the Kurdyka-Łojasiewicz (KL) inequality [4, 8, 32, 33], then the ADMM algorithm* (8) *starting from any initialization $(\mathbf{x}^0, y^0, w^0)$ globally converges to a unique stationary point $(\mathbf{x}^*, y^*, w^*)$ of the augmented Lagrangian $\mathcal{L}_\beta(\mathbf{x}, y; w)$.*

4.2 The ADMM algorithm for solving (5)

Taking advantage of the convergence result of the ADMM algorithm, we apply it to solve the nonconvex variational model in (5) for simultaneously denoising and deblurring the degraded images corrupted by Cauchy noise. From now on we proceed in discrete terms, but, for the sake of simplicity, we keep the notation from the continuous context. We assume that the discrete image domain Ω contains $n \times n$ pixels. The discrete minimization nonconvex model of (5) is formulated as follows:

$$\min_{u\in\mathbb{R}^{n^2}} \|\nabla u\|_1 + \frac{\lambda}{2}\left\langle \log\left(\gamma^2 + (Ku - f)^2\right), \mathbf{1}\right\rangle, \tag{9}$$

[1] Note that the full column rank assumption can be weakened by the following assumption: for the general matrix $\mathbf{A}$ and B, there exists two Lipschitz continuous maps such that $\mathcal{H}_1(u) = \arg\min_x\{\mathcal{F}(\mathbf{x}) : \mathbf{A}\mathbf{x} = u\}$ and $\mathcal{H}_2(v) = \arg\min_y\{\mathcal{G}(y) : By = v\}$.

[2] A function $h : \mathbb{R}^N \to \mathbb{R}$ is restricted prox-regular, if for any sufficiently large $M \in \mathbb{R}_+$ and any bounded set $T \subset \mathbb{R}^N$, there exists $\tau > 0$ such that

$$h(y) + \frac{\tau}{2}\|x - y\|^2 \geq h(x) + \langle d, y - x\rangle, \text{ for all } x, y \in T \setminus S_M, d \in \partial h(x), \|d\| \leq M$$

where $S_M := \{d \in dom(\partial h) : \|d\| > M \text{ for all } d \in \partial h\}$ is the exclusion.

where $f \in \mathbb{R}^{n^2}$ is obtained by stacking the columns of the corresponding $n \times n$ gray-scale image, and $K \in \mathbb{R}^{n^2 \times n^2}$. The TV regularization $\|\nabla u\|_1$ is given as,

$$\|\nabla u\|_1 = \sum_{i=1}^{n^2} \sqrt{(\nabla_x u)_i^2 + (\nabla_y u)_i^2},$$

where $\nabla_x \in \mathbb{R}^{n^2 \times n^2}$ and $\nabla_y \in \mathbb{R}^{n^2 \times n^2}$ are the discrete first order forward differences in the x- and y-direction under the periodic boundary condition, respectively. The discrete gradient of u, ∇u, is defined as $\nabla u = [(\nabla_x u)^\top, (\nabla_y u)^\top]^\top \in \mathbb{R}^{2n^2}$.

Based on the framework of the ADMM algorithm, we introduce a new auxiliary variable $v \in \mathbb{R}^{n^2}$ and obtain the following constrained nonconvex minimization problem:

$$\begin{aligned} \min_{u,v \in \mathbb{R}^{n^2}} \ & \|\nabla u\|_1 + \frac{\lambda}{2} \left\langle \log\left(\gamma^2 + (v-f)^2\right), \mathbf{1} \right\rangle \\ \text{s.t. } & Ku = v. \end{aligned} \tag{10}$$

Let $w \in \mathbb{R}^{n^2}$ be the Lagrangian multiplier for the constraint $Ku = v$. Then we have the corresponding augmented Lagrangian function,

$$\mathcal{L}_\beta(u, v, w) = \|\nabla u\|_1 + \frac{\lambda}{2} \left\langle \log\left(\gamma^2 + (v-f)^2\right), \mathbf{1} \right\rangle + \langle w, Ku - v \rangle + \frac{\beta}{2} \|Ku - v\|_2^2,$$

where $\beta > 0$ is a penalty parameter that controls the convergence speed of the ADMM algorithm. The whole algorithm for restoring the blurred images corrupted by Cauchy noise is given in Algorithm 1.

Algorithm 1 ADMM algorithm for solving (9)

1: Initialize u^0, v^0, w^0; set λ, β.

2: For $k = 1, 2, \ldots$, calculate iteratively u^{k+1}, v^{k+1},w^{k+1}.

$$u^{k+1} \in \arg\min_u \|\nabla u\|_1 + \frac{\beta}{2} \left\| Ku - v^k + \frac{w^k}{\beta} \right\|_2^2 \tag{11}$$

$$v^{k+1} = \arg\min_v \frac{\lambda}{2} \left\langle \log\left(\gamma^2 + (v-f)^2\right), \mathbf{1} \right\rangle + \frac{\beta}{2} \left\| Ku^{k+1} - v + \frac{w^k}{\beta} \right\|_2^2 \tag{12}$$

$$w^{k+1} = w^k + \beta(Ku^{k+1} - v^{k+1}) \tag{13}$$

3: If u^{k+1} satisfies the stopping criteria, it returns u^{k+1} and stops.

In Algorithm 1, the main computations are in the steps to solve the minimization problems in (11) and (12). The u-subproblem (11) can be efficiently solved by many methods, for instance the dual algorithm [10], the split-Bregman algorithm [9, 22, 43, 50], the primal-dual algorithm [11, 18] ,the infeasible primal-dual algorithm of semi-smooth Newton-type [25], and the ADMM algorithm [14, 24, 52]. Here, we apply the dual algorithm proposed in [10]. Since there exists the second order derivative for the objective function in (12), we can utilize the Newton method to solve it efficiently. Inspired by [48], as a special case of (7), we have the following convergence result for Algorithm 1. In addition, taking some specific properties of the variational model (9) into account, we will give a relatively simple proof.

Theorem 4.2. *Let (u^0, v^0, w^0) be the initial value and $\{(u^k, v^k, w^k)\}$ be the iterative sequence generated by Algorithm 1. Then, if $\beta > \frac{\lambda + \sqrt{\lambda^2 + 8\lambda\gamma^2}}{2\gamma^2}$ and K has full column rank, the sequence $\{(u^k, v^k, w^k)\}$ converges globally to the unique limit point (u^*, v^*, w^*). Moreover, the unique limit point is a stationary point of $\mathcal{L}_\beta$.*

In order to prove Theorem 4.2, based on the model in (7) we define the following functions:

$$\begin{aligned} \mathcal{F} : \mathbb{R}^{n^2} \to \mathbb{R}, \qquad \mathcal{F}(u) &= \|\nabla u\|_1 \\ \mathcal{G} : \mathbb{R}^{n^2} \to \mathbb{R}, \qquad \mathcal{G}(v) &= \frac{\lambda}{2}\left\langle \log\left(\gamma^2 + (v-f)^2\right), \mathbf{1}\right\rangle. \end{aligned} \tag{14}$$

The feasible set is $\Omega_F = \{(u, v) \in \mathbb{R}^{n^2} \otimes \mathbb{R}^{n^2} : Ku - v = 0\}$. First, we give some useful lemmas that will be used in the main proof.

Lemma 4.1. *Algorithm 1 holds that,*

1. *for all $k \in \mathbb{N}$, $\nabla\mathcal{G}(v^k) = w^k$.*
2. *$\|w^k - w^{k+1}\| \le \frac{\lambda}{\gamma^2}\|v^k - v^{k+1}\|$.*

Proof. Substituting (13) on w^k into the first-order optimality condition of the v-subproblem on v^k: $\nabla\mathcal{G}(v^k) - w^{k-1} + \beta(v^k - Ku^k) = 0$, we have $\nabla\mathcal{G}(v^k) = w^k$ for all $k \in \mathbb{N}$.

Since $\mathcal{G}$ is smooth, we can calculate its second derivative

$$\partial^2_{v_i}\mathcal{G} = \lambda\frac{\gamma^2 - (v_i - f)^2}{(\gamma^2 + (v_i - f)^2)^2} \le \frac{\lambda}{\gamma^2},$$

and $L_{\nabla\mathcal{G}} = \frac{\lambda}{\gamma^2}$ is Lipschitz constant for $\nabla\mathcal{G}$. Then, we obtain

$$\|w^k - w^{k+1}\| = \|\nabla\mathcal{G}(v^k) - \nabla\mathcal{G}(v^{k+1})\| \le L_{\nabla\mathcal{G}}\|v^k - v^{k+1}\|.$$

□

Lemma 4.2. *Let $\{(u^k, v^k, w^k)\}$ be the iterative sequence generated by Algorithm 1. If $\beta > \frac{\lambda + \sqrt{\lambda^2 + 8\lambda\gamma^2}}{2\gamma^2}$, the iterative sequence $\{(u^k, v^k, w^k)\}$ satisfies:*

1. *$\mathcal{L}_\beta(u^k, v^k, w^k)$ is lower bounded and nonincreasing for all $k \in \mathbb{N}$.*
2. *$\{(u^k, v^k, w^k)\}$ is bounded.*

Proof. According to the optimality condition of the u-subproblem (11), we define

$$d^{k+1} := -(K^\top w^k + \beta K^\top(Ku^{k+1} - v^k)) \in \partial\mathcal{F}(u^{k+1}). \tag{15}$$

From the equation (11) and the definition of subgradient $\mathcal{F}$, it follows

$$\begin{aligned} &\mathcal{L}_\beta(u^k, v^k, w^k) - \mathcal{L}_\beta(u^{k+1}, v^k, w^k) \\ &= \mathcal{F}(u^k) - \mathcal{F}(u^{k+1}) + \langle w^k, Ku^k - Ku^{k+1}\rangle + \frac{\beta}{2}\|Ku^k - v^k\|^2 - \frac{\beta}{2}\|Ku^{k+1} - v^k\|^2 \\ &= \mathcal{F}(u^k) - \mathcal{F}(u^{k+1}) + \langle K^\top w^k + \beta K^\top(Ku^{k+1} - v^k), u^k - u^{k+1}\rangle + \frac{\beta}{2}\|Ku^k - Ku^{k+1}\|^2 \\ &= \mathcal{F}(u^k) - \mathcal{F}(u^{k+1}) - \langle d^{k+1}, u^k - u^{k+1}\rangle + \frac{\beta}{2}\|Ku^k - Ku^{k+1}\|^2 \\ &\ge \frac{\beta}{2}\|Ku^k - Ku^{k+1}\|^2 \end{aligned} \tag{16}$$

where the second equality follows from the cosine rule: $\|b+c\|^2-\|a+c\|^2 = \|b-a\|^2+2\langle a+c, b-a\rangle$ and the last inequality follows from the fact that $\mathcal{F}(u)$ is convex.

For the updates of v^{k+1}, w^{k+1}, by the cosine rule and Lemma 4.1, we have

$$\begin{aligned}
&\mathcal{L}_\beta(u^{k+1}, v^k, w^k) - \mathcal{L}_\beta(u^{k+1}, v^{k+1}, w^{k+1}) \\
&= \mathcal{G}(v^k) - \mathcal{G}(v^{k+1}) + \langle w^{k+1}, v^{k+1} - v^k\rangle - \beta\|Ku^{k+1} - v^{k+1}\|^2 + \frac{\beta}{2}\|v^k - v^{k+1}\|^2 \\
&= \mathcal{G}(v^k) - \mathcal{G}(v^{k+1}) - \langle \nabla\mathcal{G}(v^{k+1}), v^k - v^{k+1}\rangle - \frac{1}{\beta}\|w^k - w^{k+1}\|^2 + \frac{\beta}{2}\|v^k - v^{k+1}\|^2 \\
&\geq -\frac{L_{\nabla\mathcal{G}}}{2}\|v^k - v^{k+1}\|^2 - \frac{L_{\nabla\mathcal{G}}}{\beta}\|v^k - v^{k+1}\|^2 + \frac{\beta}{2}\|v^k - v^{k+1}\|^2 \\
&= C\|v^k - v^{k+1}\|^2
\end{aligned} \tag{17}$$

where $C = \frac{\beta}{2} - \frac{L_{\nabla\mathcal{G}}}{2} - \frac{L_{\nabla\mathcal{G}}}{\beta}$. In order to make sure $C > 0$, we need that the penalty parameter β satisfies:

$$\beta > \frac{L_{\nabla\mathcal{G}} + \sqrt{L_{\nabla\mathcal{G}}^2 + 8L_{\nabla\mathcal{G}}}}{2} = \frac{\lambda + \sqrt{\lambda^2 + 8\lambda\gamma^2}}{2\gamma^2}.$$

According to (16) and (17), we have

$$\mathcal{L}_\beta(u^k, v^k, w^k) - \mathcal{L}_\beta(u^{k+1}, v^{k+1}, w^{k+1}) \geq C\|v^k - v^{k+1}\|^2 + \frac{\beta}{2}\|Ku^k - Ku^{k+1}\|^2.$$

This means that $\mathcal{L}_\beta(u^k, v^k, w^k)$ is nonincreasing for all $k \in \mathbb{N}$.

As K has full column rank, there exists $\hat{v}$ such that $Ku^k - \hat{v} = 0$. Therefore, we have

$$\mathcal{F}(u^k) + \mathcal{G}(\hat{v}) \geq \min_{u,v}\{\mathcal{F}(u) + \mathcal{G}(v) : Ku - v = 0\} > -\infty.$$

So we obtain the following result

$$\begin{aligned}
\mathcal{L}_\beta(u^k, v^k, w^k) &= \mathcal{F}(u^k) + \mathcal{G}(v^k) + \langle w^k, Ku^k - v^k\rangle + \frac{\beta}{2}\|Ku^k - v^k\|^2 \\
&= \mathcal{F}(u^k) + \mathcal{G}(v^k) + \langle \nabla\mathcal{G}(v^k), \hat{v} - v^k\rangle + \frac{\beta}{2}\|Ku^k - v^k\|^2 \\
&\geq \mathcal{F}(u^k) + \mathcal{G}(\hat{v}) - \frac{L_{\nabla\mathcal{G}}}{2}\|\hat{v} - v^k\|^2 + \frac{\beta}{2}\|Ku^k - v^k\|^2 \\
&= \mathcal{F}(u^k) + \mathcal{G}(\hat{v}) + \frac{\beta - L_{\nabla\mathcal{G}}}{2}\|Ku^k - v^k\|^2 > -\infty
\end{aligned}$$

Since $\mathcal{L}_\beta(u^k, v^k, w^k)$ is upper bounded by $\mathcal{L}_\beta(u^0, v^0, w^0)$ and obviously $\mathcal{F}(u) + \mathcal{G}(v)$ is coercive over Ω_F, we obtain that $\{u^k\}$ and $\{v^k\}$ are bounded. Based on Lemma 4.1, we have that $\{w^k\}$ is also bounded. □

Lemma 4.3. *Set $\partial\mathcal{L}(u^{k+1}, v^{k+1}, w^{k+1}) = (\partial_u\mathcal{L}, \partial_v\mathcal{L}, \partial_w\mathcal{L})$. Then there exists a constant $C_1 > 0$ such that $\|\partial\mathcal{L}(u^{k+1}, v^{k+1}, w^{k+1})\| \leq C_1\|v^k - v^{k+1}\|$.*

Proof. Because $\partial_w\mathcal{L} = Ku^{k+1} - v^{k+1} = \frac{1}{\beta}(w^{k+1} - w^k)$ and $\partial_v\mathcal{L} = w^{k+1} - w^k$, based on Lemma 4.1, we have

$$\|\partial_w\mathcal{L}\| \leq \frac{L_{\nabla\mathcal{G}}}{\beta}\|v^k - v^{k+1}\|, \tag{18}$$

$$\|\partial_v\mathcal{L}\| \leq L_{\nabla\mathcal{G}}\|v^k - v^{k+1}\|. \tag{19}$$

By the definition of the subgradient, we have

$$\begin{aligned}\partial_u \mathcal{L} &= \partial \mathcal{F}(u^{k+1}) + K^\top w^{k+1} + \beta K^\top (K u^{k+1} - v^{k+1}) \\ &= \partial \mathcal{F}(u^{k+1}) + K^\top w^k + \beta K^\top (K u^{k+1} - v^k) + K^\top (w^{k+1} - w^k) + \beta K^\top (v^k - v^{k+1}).\end{aligned} \tag{20}$$

Thus, according to the optimal condition $0 \in \partial \mathcal{F}(u^{k+1}) + K^\top w^k + \beta K^\top (K u^{k+1} - v^k)$, we have $K^\top (w^{k+1} - w^k) + \beta K^\top (v^k - v^{k+1}) \in \partial_u \mathcal{L}$. Combining (18), (19), (20) and Lemma 4.1, we obtain

$$\begin{aligned}\|\partial \mathcal{L}(u^{k+1}, v^{k+1}, w^{k+1})\| &\le \left(L_{\nabla \mathcal{G}} \left(1 + \frac{1}{\beta} + \|K\| \right) + \beta \|K\| \right) \|v^k - v^{k+1}\| \\ &= C_1 \|v^k - v^{k+1}\|\end{aligned}$$

where $C_1 = \left(L_{\nabla \mathcal{G}} (1 + \frac{1}{\beta} + \|K\|) + \beta \|K\| \right)$. □

In the end of this section, we give the proof for the convergence theorem.

Proof. (of Theorem 4.2)
As K has full column rank, the feasible set Ω_F is nonempty. By Lemma 4.2, the iterative sequence $\{(u^k, v^k, w^k)\}$ is bounded, so there exists a convergent subsequence $\{(u^{n_k}, v^{n_k}, w^{n_k})\}$, i.e., $(u^{n_k}, v^{n_k}, w^{n_k})$ converges to (u^*, v^*, w^*) as k goes to infinity. Due to the nonincreasing and lower-bounded $\mathcal{L}_\beta(u^k, v^k, w^k)$, we have $\|K(u^k - u^{k+1})\| \to 0$ and $\|v^k - v^{k+1}\| \to 0$ as $k \to \infty$. According to Lemma 4.3, there exists $p^k \in \partial \mathcal{L}_\beta(u^k, v^k, w^k)$ such that $\|p^k\| \to 0$. Further, this leads to $\|p^{n_k}\| \to 0$ as $k \to \infty$. Based on the definition of the general subgradient [39], we obtain that $0 \in \partial \mathcal{L}_\beta(u^*, v^*, w^*)$, i.e., (u^*, v^*, w^*) is a stationary point.

Referring to [47, 51], the function $\mathcal{F}(u)$ is semi-algebraic and $\mathcal{G}(v)$ is a real analytic function. Thus, we conclude that $\mathcal{L}_\beta$ satisfies the KL inequality [8]. Then, as in the proof of Theorem 2.9 in [5], we can deduce that the iterative sequence $\{(u^k, v^k, w^k)\}$ is globally convergent to (u^*, v^*, w^*).
□

Remark 1. *In the Theorem 4.2 we need that K has full column rank. Since K is a blurring matrix in our problem, this requirement does not impose serious restriction. Especially under the periodic boundary condition, this condition is satisfied.*

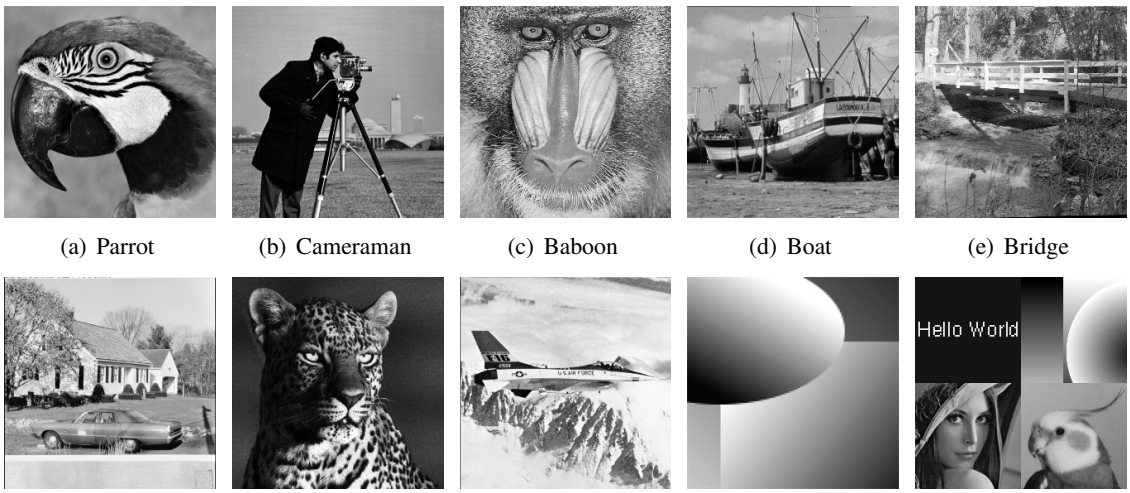
5 Numerical Experiments

In this section, we present several numerical experiments to demonstrate the performance of the proposed method for restoring the blurred images corrupted by Cauchy noise. Here, we use ten 8-bit 256-by-256 gray-scale test images, see Figure 2. All numerical results are performed under Windows 10 and Matlab Version 7.10 (R2012a) running on a Lenovo laptop with a 1.7GHz Intel Core CPU and 4G RAM memory.

In order to evaluate the performance of the proposed method, we utilize the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) [49], which are defined as

$$\text{PSNR} = 20 \log 10 \left(\frac{255n}{\|\tilde{u} - u\|_2} \right), \quad \text{SSIM} = \frac{2\mu_{\tilde{u}}\mu_u(2\sigma + c_2)}{(\mu_{\tilde{u}}^2 + \mu_u^2 + c_1)(\sigma_{\tilde{u}}^2 + \sigma_u^2 + c_2)},$$

where $\tilde{u}$ is the restored image, u is the original image, $\mu_{\tilde{u}}$ and μ_u denote their means, $\sigma_{\tilde{u}}^2$ and σ_u^2 represent their variances, σ is the covariance of $\tilde{u}$ and u, and $c_1, c_2 > 0$ are constants. The values of PSNR are satisfied for the human subjective sensation, which the higher PSNR value implies the



(a) Parrot (b) Cameraman (c) Baboon (d) Boat (e) Bridge

(f) House (g) Leopard (h) Plane (i) Test (j) Montage

Figure 2. Original images.

better quality of restored images. The SSIM conforms with the quality perception of the human visual system (HVS). If the SSIM value is closer to 1, the characteristic (edges and textures) of restored images is more similar to the original image.

In our method, we set the stopping condition based on the following relative improvement inequality,

$$\left|\frac{E(u^{k+1}) - E(u^k)}{E(u^{k+1})}\right| < \epsilon \tag{21}$$

where E is the objective function in (9) and $\epsilon = 5 \times 10^{-5}$. In addition, since the regularization parameter λ balances the trade-off of the good fit to f and the smoothness from the TV, we manually tune it in order to obtain the highest PSNR values of the restored images. The selection method with respect to λ is out of the scope in this paper. The parameter β in Algorithm 1 controls the convergent speed. Based on Theorem 4.2, we round $\beta > \frac{\lambda + \sqrt{\lambda^2 + 8\lambda\gamma^2}}{2\gamma^2}$ up to the nearest value with two digits after the decimal point as β. In addition, we set the iteration number for the Newton method in solving v-subproblem as 3. The iteration number for solving u-subproblem equals 5 for the denoising case and 10 for the simultaneous deblurring and denoising.

5.1 Different initializations

Since the model (9) is nonconvex, although we are able to prove that the ADMM algorithm converges globally to a unique stationary point with given starting point (u^0, v^0, w^0), the local minimizers that we obtained still depends on the initializations. To study the influence of initializations and find the one which provides better restoration, in this section we test on three different choices of u^0 on denoising case:

$$\begin{aligned}(I)\ &: u^0 = \max(0, \min(f, 255));\\ (II)\ &: u^0 = medfile2(f);\\ (III)\ &: u^0 = f,\end{aligned}$$

Table 1. PSNR and SSIM for the test images "Parrot" and "Cameraman" with different initial values.

Noise	Condition	Parrot		Cameraman	
		PSNR	SSIM	PSNR	SSIM
$\gamma = 5$	I	29.06	0.8729	28.72	0.8520
	II	27.83	0.8505	26.31	0.8500
	III	24.88	0.7730	23.75	0.7275
$\gamma = 10$	I	27.12	0.8268	26.67	0.7949
	II	26.68	0.8218	25.60	0.8093
	III	22.87	0.6895	22.43	0.6653

(a) (b) (c) (d)

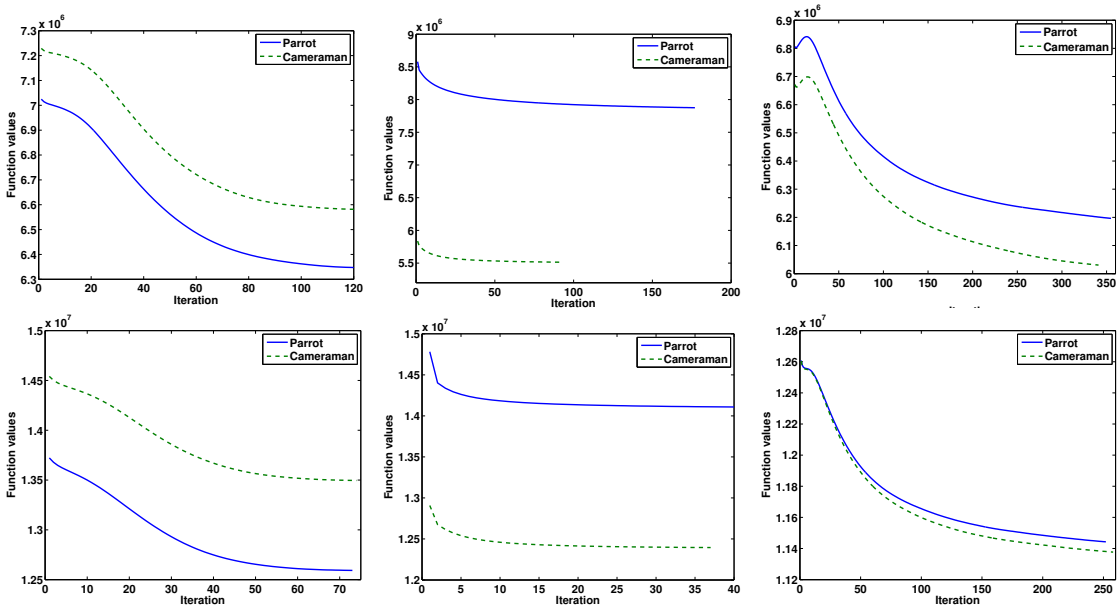
Figure 3. Comparison of different initial values for removing Cauchy noise in the image "Parrot", with $\gamma = 5$ (in the 1st row) and 10 (in the 2nd row). (a) noisy images, (b) restored images of (I), (c) restored images of (II), (d) restored images of (III).

where $medfile2(f)$ denotes the result from the median filter with window size 3. Note that due to the impulsive features of Cauchy noise, the median filter usually provides fairly good results. In addition, based on Theorem 3.2 with u^0 in case (III) we will obtain the unique solution in Ω_U.

In Table 1, we list the values of PSNR and SSIM by using different initializations in our method for the test images "Parrot" and "Cameraman" with the noise level $\gamma = 5$ and 10. It is obvious that the values of both PSNR and SSIM are highest in the case (I), and are lowest in the case (III). Hence, the unique solution in Ω_U is not a satisfactory local minimizer.

In Figure 3 we show the restored "Parrot" images in order to compare the visual performance by applying different initial values. In Figure 3 (d), we see that there is still some noise left in restored images. Compared with the results from (II), the ones from (I) include clearer features with less noise, especially in the region around the eye and black stripes of "Parrot". Hence, we choose (I) as the initial value in the following numerical experiments.

Theorem 4.2 demonstrate that with given initial value Algorithm 1 converges globally to a unique stationary point. In Figure 4, we show the plots of the objective function values in (9) versus the number of iteration in order to observe the convergence of our method. It's clear that the objective function value keeps decreasing along the iterations. Furthermore, our method basically



(a) I (b) II (c) III

Figure 4. Plots of the objective function values versus iterations for the noisy images "Parrot" and "Cameraman" with $\gamma = 5$ (in the 1st row) and 10 (in the 2nd row).

converges very fast except in case (III), which does not provide good restorations.

5.2 Comparisons of image deblurring and denoising

In order to show the superior performance of the proposed method, we compare our method with the other two well-known methods: the median filter (matlab function '*medfilt2*') with window size 3 and the convex variational method proposed in [42] ("conRE" for short). For fair comparison, we use the same stopping rule in the convex variational method, and adjust the two parameters in the model in turn to obtain the highest PSNR values.

First, we compare the three methods for Cauchy noise removal, i.e., by setting K as identity operator. The noisy images are obtained by the degradation $f = u + \gamma\frac{\eta_1}{\eta_2}$, where η_1 and η_2 are independent random variables following Gaussian distribution with mean 0 and variance 1. Table 2 and 3 list the values of PSNR and SSIM for the restored images with the noise levels $\gamma = 5$ and $\gamma = 10$, respectively. Obviously, comparing with the other two variational methods, the median filter provides worst results with respect to the values of PSNR and SSIM. We always obtain the highest PSNR values by our method. Especially with lower noise level (γ=5), the PSNR values from our method are around 1dB higher than the convex method in [42]. Furthermore, in most cases, the SSIM values from our methods are higher than others.

In Figure 5 and 6, we show the results from different methods for removing Cauchy noise in images "Parrot", "Cameraman", "Baboon", "Boat" and "Plane". Although the median filter effectively removes Cauchy noise, it also oversmooths the edges and losses many details. It is obvious that two variational methods outperform the median filter. Comparing with the convex variational method, our method is able to preserve more details and remove most Cauchy noise. To further illustrate the performance of our method, we show the zoomed regions of the restored images "Parrot", "Baboon" and "Boat" in Figure 7 and 8, where we can clearly see the difference among the results from the three methods, e.g., the stripes around the eye in "Parrot", the nose and

Table 2. PSNR and SSIM for the noisy images and the restored images by applying different methods ($\gamma = 5$).

Image	Noisy	PSNR			SSIM		
		Median	conRE	Ours	Median	conRE	Ours
Parrot	19.20	27.18	27.19	**29.06**	0.8341	0.8465	**0.8729**
Cameraman	18.98	25.94	26.51	**28.72**	0.7996	0.8225	**0.8520**
Baboon	17.74	19.18	21.18	**22.56**	0.5069	0.7178	**0.7781**
Boat	18.01	25.94	27.03	**27.94**	0.7779	0.8165	**0.8541**
Bridge	19.13	22.63	24.32	**25.25**	0.6312	0.7857	**0.8097**
House	17.94	24.06	25.25	**26.26**	0.7510	0.7774	**0.8308**
Leopard	19.07	25.34	26.54	**27.51**	0.7787	0.7861	**0.8309**
Plane	17.37	25.09	25.83	**27.23**	0.8235	0.8354	**0.8611**
Test	19.19	34.79	39.55	**40.77**	0.8922	0.9726	**0.9864**
Montage	19.14	27.52	27.94	**31.08**	0.8772	**0.9180**	0.9135

Table 3. PSNR and SSIM for the noisy images and the restored images by applying different methods ($\gamma = 10$).

Image	Noisy	PSNR			SSIM		
		Median	conRE	Ours	Median	conRE	Ours
Parrot	16.35	25.51	26.74	**27.12**	0.7254	0.8202	**0.8268**
Cameraman	16.06	24.68	25.68	**26.67**	0.6801	0.7896	**0.7981**
Baboon	14.87	18.79	20.27	**20.96**	0.4671	0.6336	**0.6634**
Boat	15.11	24.39	25.71	**25.79**	0.6843	0.7710	**0.7769**
Bridge	16.30	21.94	23.37	**23.58**	0.5870	**0.7123**	0.7033
House	15.01	22.91	24.24	**24.37**	0.6631	0.7465	**0.7582**
Leopard	16.16	24.16	25.40	**25.60**	0.6981	0.7649	**0.7800**
Plane	14.49	23.64	24.85	**25.25**	0.7104	0.8085	**0.8161**
Test	16.29	30.45	37.38	**38.01**	0.7078	0.9607	**0.9793**
Montage	16.27	26.10	27.16	**28.89**	0.7451	**0.8922**	0.8850

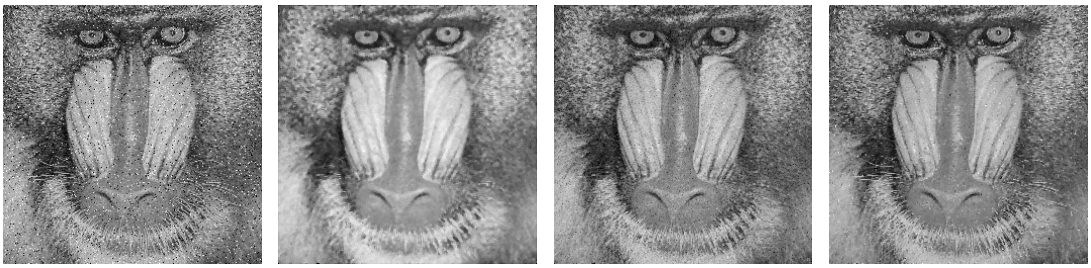
whiskers of "Baboon", and the ropes and iron pillars of "Boat".

In the following experiments, we compare the three methods on recovering the blurred images corrupted by Cauchy noise. Here, we consider the Gaussian blurring operator with size 7 and standard deviation 3, and the out-of-focus blurring operator with size 5. Further, Cauchy noise with $\gamma = 5$ is added into the blurry images. Table 4 and 5 list the values of the PSNR and SSIM by applying different methods for the images "Parrot", "Cameraman", "Plane" and "Test". Figure 9 and 10 show the restored images for deblurring and denoising the degraded images.

From Table 4 and 5, we find that our method provides the largest values on both PSNR and SSIM. Comparing with the convex variational method, our method can improve at least 0.36dB on PSNR values. In Figure 9 and 10, it is clear that the restored images by the median filter are oversmoothing, since the median filter cannot deal with deblurring. The convex variational method can recover edges and textures, but some noise is still not removed. However, our method not only preserves the fine features but also effectively removes Cauchy noise, which can be clearly seen from zoomed regions shown in Figure 11.

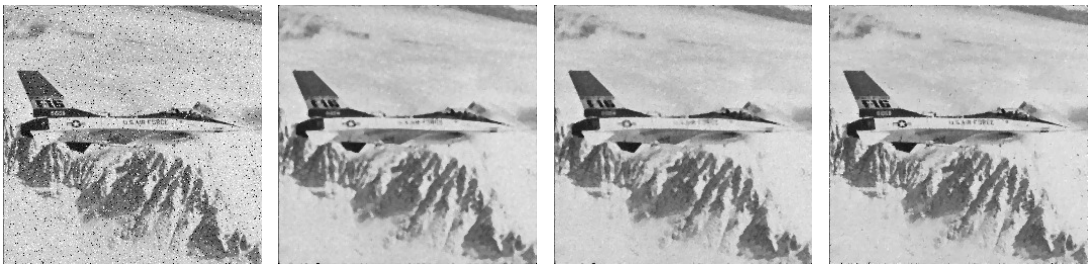
(a) Noisy image:19.20 (b) Median:27.18 (c) conRE:27.91 (d) Ours:29.06

(e) Noisy image:18.98 (f) Median:25.94 (g) conRE:26.51 (h) Ours:28.72



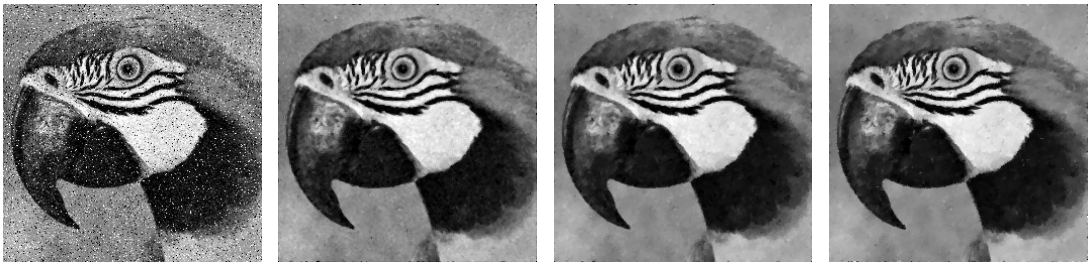
(i) Noisy image:17.74 (j) Median:19.18 (k) conRE:21.18 (l) Ours:22.56

(m) Noisy image:18.01 (n) Median:25.94 (o) conRE:27.03 (p) Ours:27.94

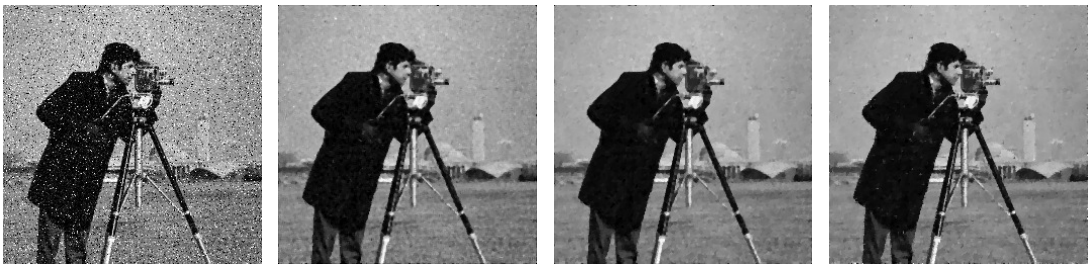


(q) Noisy image:17.37 (r) Median:25.09 (s) conRE:25.83 (t) Ours:27.25

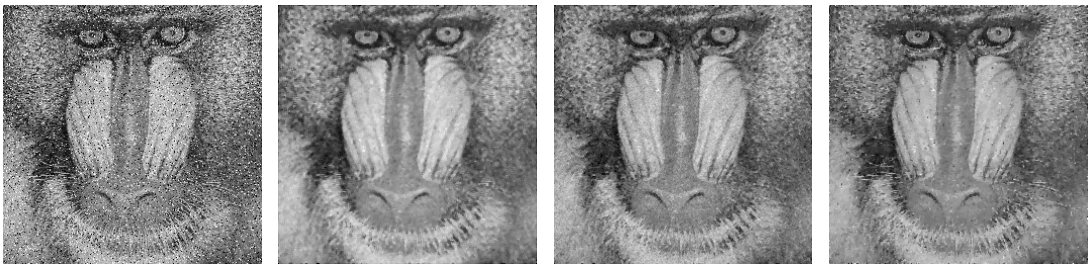
Figure 5. Comparison of different methods for removing Cauchy noise, with $\gamma = 5$.



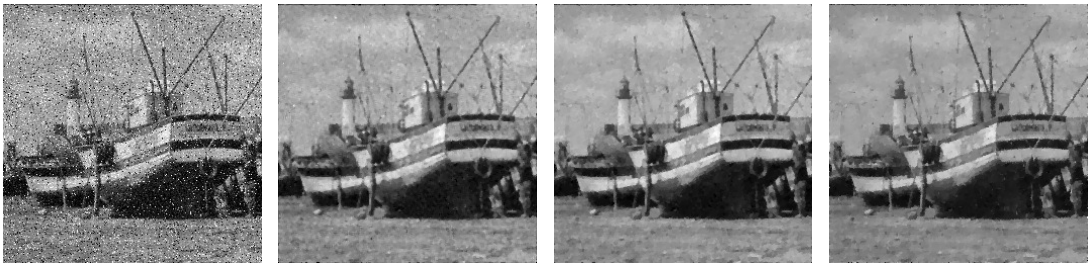
(a) Noisy image:16.35 (b) Median:25.51 (c) conRE:26.74 (d) Ours:27.12



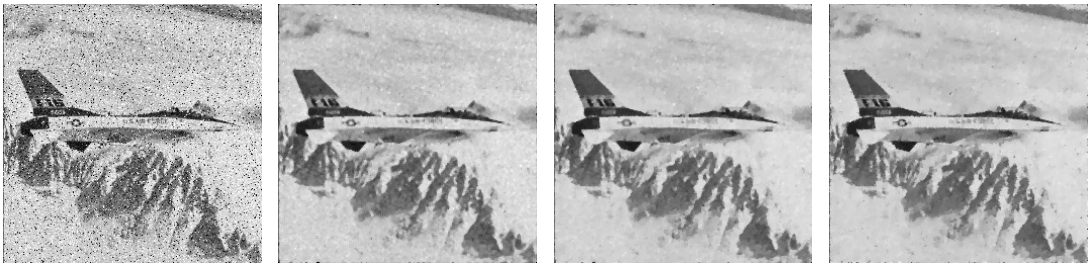
(e) Noisy image:16.06 (f) Median:24.68 (g) conRE:25.68 (h) Ours:26.67



(i) Noisy image:14.87 (j) Median:18.79 (k) conRE:20.27 (l) Ours:20.96

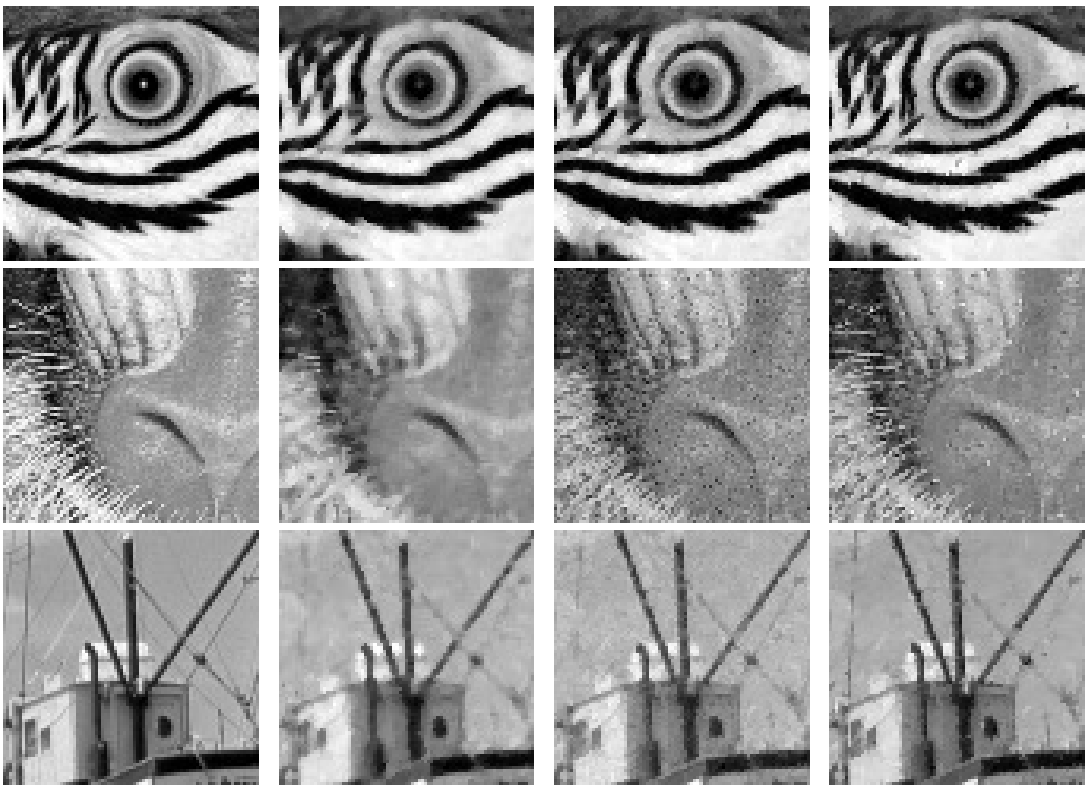


(m) Noisy image:15.11 (n) Median:24.39 (o) conRE:25.71 (p) Ours:25.79



(q) Noisy image:14.49 (r) Median:23.64 (s) conRE:24.85 (t) Ours:25.25

Figure 6. Comparison of different methods for removing Cauchy noise, with $\gamma = 10$.



(a) (b) (c) (d)

Figure 7. Zoomed version of the restored images in Figure 5. (a) original images; (b) the median filter; (c) the "conRe" model; (d) our method.

6 Conclusion

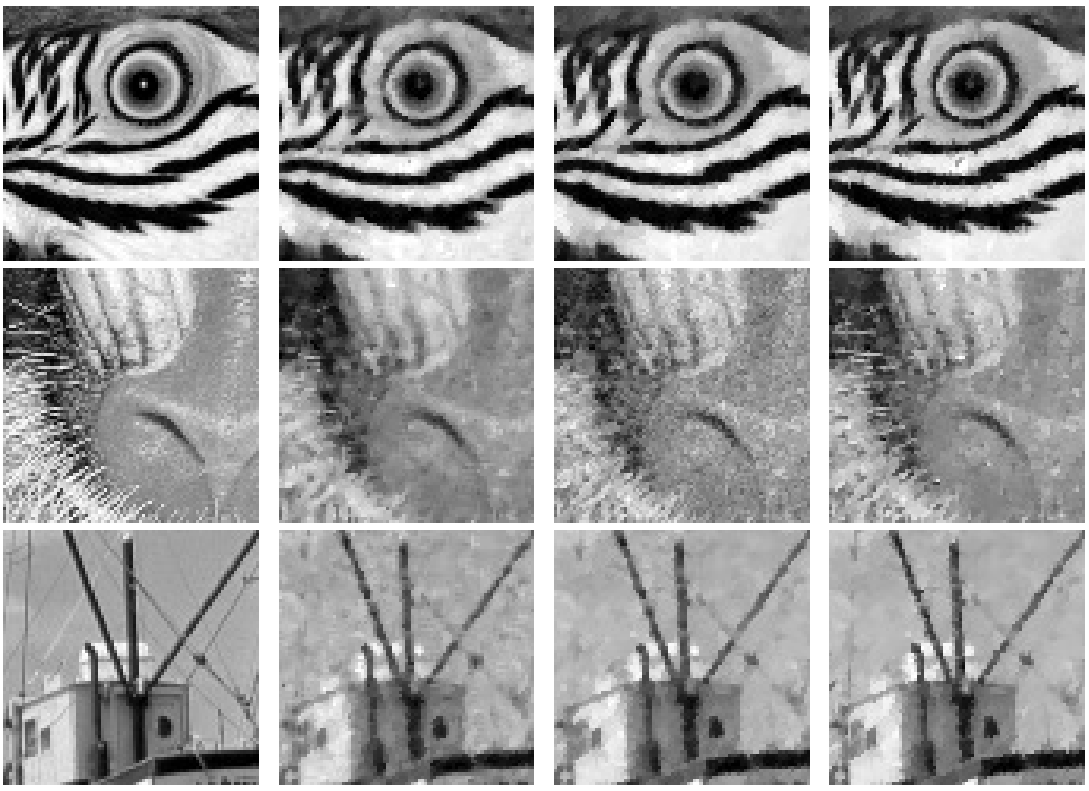
In this paper, we review and analyze the statistic properties of Cauchy distribution by comparing with Gaussian distribution and Laplace distribution. According to the MAP estimator, we illuminate the nonconvex variational model for restoring images degraded by blur and Cauchy noise. Taking advantage of the recent results in [48], the alternating direction method of multiplier (ADMM) algorithm is applied to solve the nonconvex variational optimization problem, and it is proved with global convergence. Numerical experiments show that the proposed method outperforms another two well-known methods with respect to the qualitative and quantitative comparisons.

Acknowledgements

We would like to thank Federica Sciacchitano for providing the codes of the method in [42].

References

[1] A. Achim and E. E. Kuruoğlu, *Image denoising using bivariate α-stable distributions in the complex wavelet domain*, IEEE Signal Processing Letters, 12 (2005), pp. 17–20.



(a) (b) (c) (d)

Figure 8. Zoomed version of the restored images in Figure 6. (a) original images; (b) the median filter; (c) the "conRe" model; (d) our method.

[2] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, vol. 254, Oxford university press, 2000.

[3] B. C. Arnold and R. J. Beaver, *The skew-Cauchy distribution*, Statistics & probability letters, 49 (2000), pp. 285–290.

[4] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Lojasiewicz inequality*, Mathematics of Operations Research, 35 (2010), pp. 438–457.

[5] H. Attouch, J. Bolte, and B. F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized gauss–seidel methods*, Mathematical Programming, 137 (2013), pp. 91–129.

[6] N. Balakrishnan and V. B. Nevzorov, *A primer on statistical distributions*, John Wiley & Sons, 2004.

[7] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: numerical methods*, (1997).

[8] J. Bolte, A. Daniilidis, and A. Lewis, *The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM Journal on Optimization, 17 (2007), pp. 1205–1223.

Table 4. PSNR and SSIM for the images degraded by Gaussian blur and Cauchy noise ($\gamma = 5$) and the restored images by different methods.

Image	Noisy	PSNR			SSIM		
		Median	conRE	Ours	Median	conRE	Ours
Parrot	17.16	21.13	23.73	**24.45**	0.6698	0.7174	**0.7649**
Cameraman	17.17	21.61	23.51	**24.10**	0.6397	0.6846	**0.7490**
Plane	15.74	20.52	22.11	**22.57**	0.6508	0.6666	**0.7521**
Test	18.86	29.60	32.31	**36.71**	0.8713	0.8263	**0.9647**

Table 5. PSNR and SSIM for the images degraded by the out-of-focus blur and Cauchy noise ($\gamma = 5$) and the restored images by different methods.

Image	Noisy	PSNR			SSIM		
		Median	conRE	Ours	Median	conRE	Ours
Parrot	17.59	22.26	25.02	**25.45**	0.7157	0.7496	**0.8005**
Cameraman	17.42	22.36	24.62	**25.08**	0.6777	0.7125	**0.7735**
Plane	15.98	21.31	23.41	**23.77**	0.6956	0.7210	**0.7854**
Test	18.94	30.61	33.68	**36.97**	0.8784	0.8553	**0.9560**

[9] J.-F. Cai, S. Osher, and Z. Shen, *Split Bregman methods and frame based image restoration*, Multiscale modeling & simulation, 8 (2009), pp. 337–369.

[10] A. Chambolle, *An algorithm for total variation minimization and applications*, Journal of Mathematical imaging and vision, 20 (2004), pp. 89–97.

[11] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision, 40 (2011), pp. 120–145.

[12] R. H. Chan, Y. Dong, and M. Hintermüller, *An efficient two-phase L^1-TV method for restoring blurred images with impulse noise*, IEEE Transactions on Image Processing, 19 (2010), pp. 1731–1739.

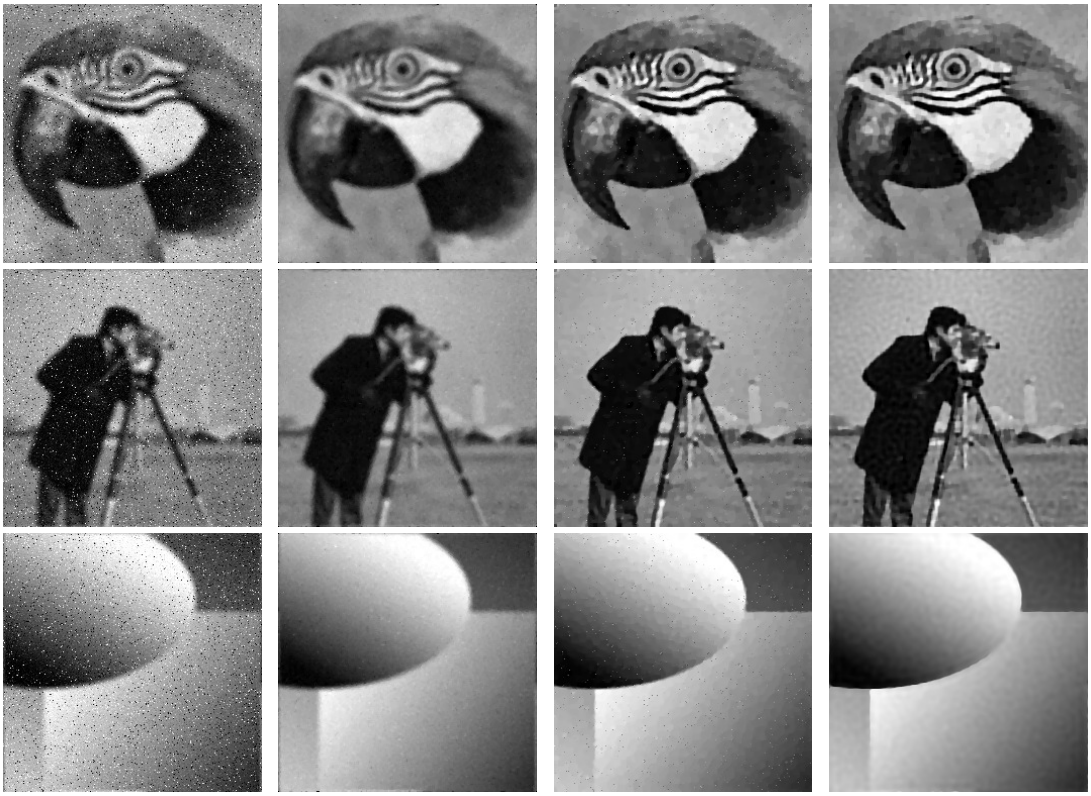
[13] Y. Chang, S. Kadaba, P. Doerschuk, and S. Gelfand, *Image restoration using recursive Markov random field models driven by Cauchy distributed noise*, IEEE Signal Processing Letters, 8 (2001), pp. 65–66.

[14] C. Chen, M. K. Ng, and X.-L. Zhao, *Alternating direction method of multipliers for nonlinear image restoration problems*, IEEE Transactions on Image Processing, 24 (2015), pp. 33–43.

[15] J. B. Copas, *On the unimodality of the likelihood for the Cauchy distribution*, Biometrika, 62 (1975), pp. 701–704.

[16] Y. Dong, M. Hintermüller, and M. Neri, *An efficient primal-dual method for L^1TV image restoration*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 1168–1189.

[17] Y. Dong and T. Zeng, *A convex variational model for restoring blurred images with multiplicative noise*, SIAM Journal on Imaging Sciences, 6 (2013), pp. 1598–1625.



(a) (b) (c) (d)

Figure 9. Comparison of the restored results by applying different methods for deblurring and denoising the images degraded by a Gaussian blur $(G, 7, 3)$ and Cauchy noise ($\gamma = 5$). (a) Degraded images; (b) the median filter; (c) the "conRe" model; (d) our method.

[18] E. Esser, X. Zhang, and T. F. Chan, *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, SIAM Journal on Imaging Sciences, 3 (2010), pp. 1015–1046.

[19] W. Feller, *An introduction to probability theory and its applications*, vol. 2, John Wiley & Sons, 2008.

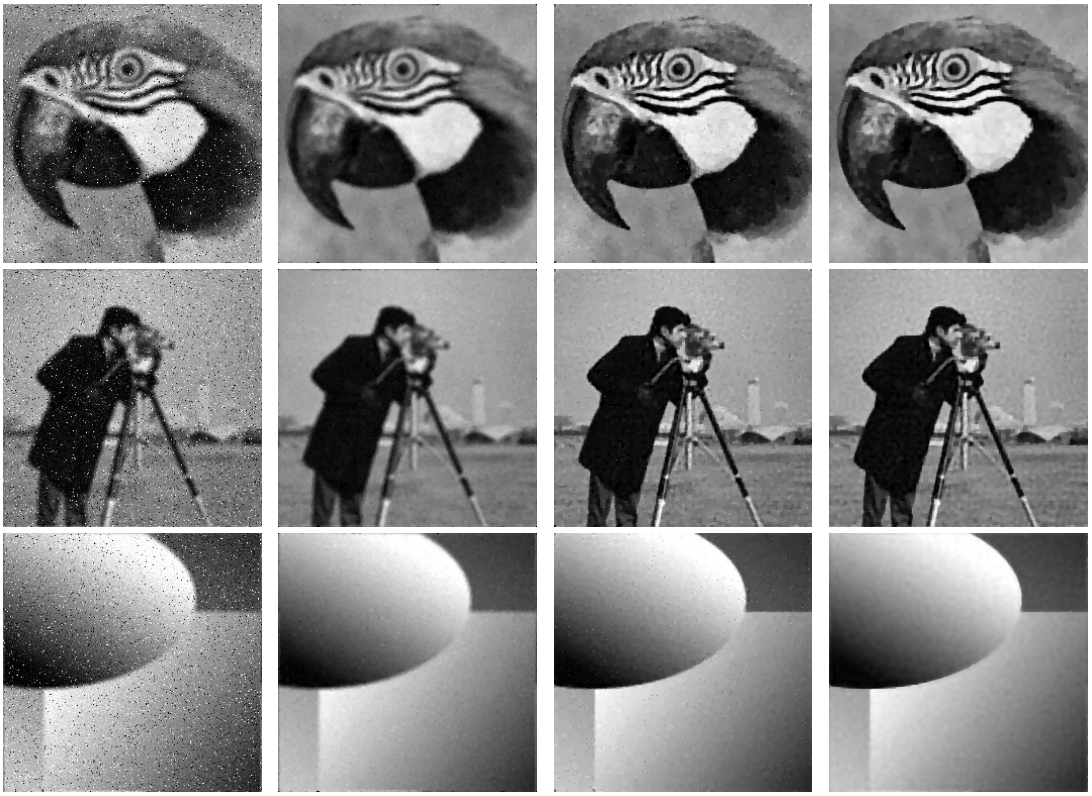
[20] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite element approximation*, Computers and Mathematics with Applications, 2 (1976), pp. 17–40.

[21] E. Giusti, *Minimal surfaces and functions of bounded variation*, no. 80, Springer Science & Business Media, 1984.

[22] T. Goldstein and S. Osher, *The split Bregman method for L1-regularized problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 323–343.

[23] M. L. Goncalves, J. G. Melo, and R. D. Monteiro, *Improved pointwise iteration-complexity of a regularized ADMM and of a regularized non-Euclidean HPE framework*, arXiv preprint arXiv:1601.01140, (2016).

[24] B. He and X. Yuan, *On the O(1/n) convergence rate of the Douglas-Rachford alternating direction method*, SIAM Journal on Numerical Analysis, 50 (2012), pp. 700–709.



(a) (b) (c) (d)

Figure 10. Comparison of the restored results by applying different methods for deblurring and denoising the images degraded by the out-of-focus blur $(A, 5)$ and Cauchy noise ($\gamma = 5$). (a) Degraded images; (b) the median filter; (c) the "conRe" model; (d) our method.

[25] M. Hintermüller and G. Stadler, *An infeasible primal-dual algorithm for total bounded variation-based inf-convolution-type image restoration*, SIAM Journal on Scientific Computing, 28 (2006), pp. 1–23.

[26] M. Idan and J. L. Speyer, *Cauchy estimation for linear scalar systems*, IEEE Transactions on Automatic Control, 55 (2010), pp. 1329–1342.

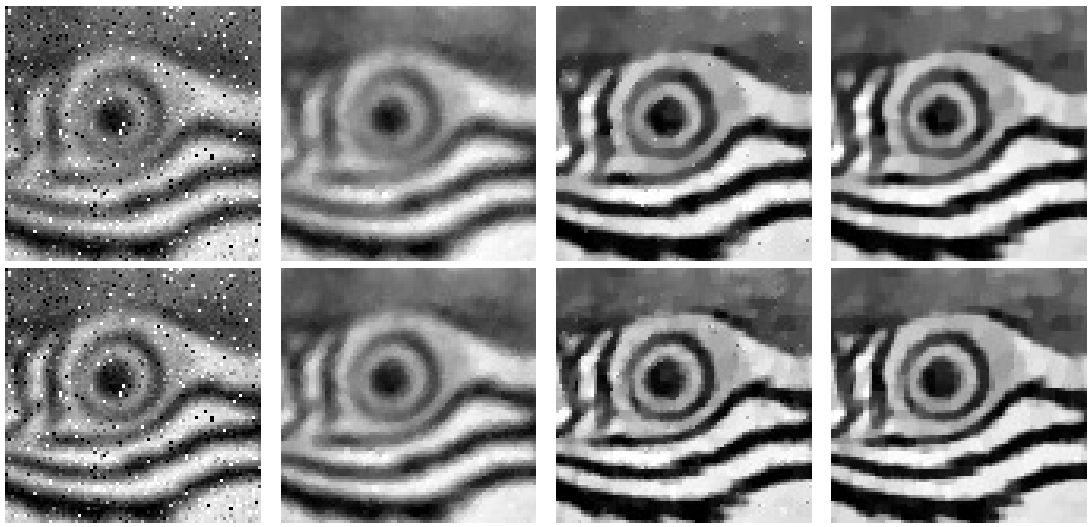
[27] T. Jeong, H. Woo, and S. Yun, *Frame-based Poisson image restoration using a proximal linearized alternating direction method*, Inverse Problems, 29 (2013), p. 075007.

[28] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous unltivariate distributions*, vol. 1, New York: John Wiley, 1994.

[29] J. T. Kent and D. E. Tyler, *Maximum likelihood estimation for the wrapped Cauchy distribution*, Journal of Applied Statistics, 15 (1988), pp. 247–254.

[30] F. B. Knight, *A characterization of the Cauchy type*, Proceedings of the American Mathematical Society, 55 (1976), pp. 130–135.

[31] E. E. Kuruoğlu, W. J. Fitzgerald, and P. J. Rayner, *Near optimal detection of signals in impulsive noise modeled with a symmetric alpha-stable distribution*, IEEE Communications Letters, 2 (1998), pp. 282–284.



(a) (b) (c) (d)

Figure 11. Zoomed version of the restored results for the image "Parrot" degraded by the Gaussian blur (in the 1st row) and the out-of-focus blur (in the 2nd row), respectively. (a) Degraded images; (b) the median filter; (c) the "conRe" model; (d) our method.

[32] S. Łojasiewicz, *Une propriété topologique des sous-ensembles analytiques réels.* Equ. Derivees partielles, Paris 1962, Colloques internat. Centre nat. Rech. sci. 117, 87-89 (1963)., 1963.

[33] S. Łojasiewicz, *Sur la géométrie semi- et sous- analytique*, Annales de l'institut Fourier, 43 (1993), pp. 1575–1595.

[34] A. Loza, D. Bull, N. Canagarajah, and A. Achim, *Non-Gaussian model-based fusion of noisy images in the wavelet domain*, Computer Vision and Image Understanding, 114 (2010), pp. 54–65.

[35] L. Ma, L. Moisan, J. Yu, and T. Zeng, *A dictionary learning approach for Poisson image deblurring*, IEEE Transactions on Medical Imaging, 32 (2013), pp. 1277–1289.

[36] M. Nikolova, *A variational approach to remove outliers and impulse noise*, Journal of Mathematical Imaging and Vision, 20 (2004), pp. 99–120.

[37] J. Nolan, *Stable distributions - models for heavy tailed data*, Birkhäuser Boston, Cambridge, MA, to appear (Chapter 1 available online from http://academic2.american.edu/˜jpnolan).

[38] J. P. Nolan, *Numerical calculation of stable densities and distribution functions*, Communications in statistics. Stochastic models, 13 (1997), pp. 759–774.

[39] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*, vol. 317, Springer Science & Business Media, 2009.

[40] G. Samoradnitsky and M. S. Taqqu, *Stable non-Gaussian random processes: stochastic models with infinite variance*, vol. 1, CRC Press, 1994.

[41] A. Sawatzky, C. Brune, T. Kösters, F. Wübbeling, and M. Burger, *EM-TV methods for inverse problems with Poisson noise*, in Level Set and PDE Based Reconstruction Methods in Imaging, Springer, 2013, pp. 71–142.

[42] F. Sciacchitano, Y. Dong, and T. Zeng, *Variational approach for restoring blurred images with Cauchy noise*, SIAM Journal on Imaging Sciences, 8 (2015), pp. 1894–1922.

[43] S. Setzer, *Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage*, in Scale space and variational methods in computer vision, Springer, 2009, pp. 464–476.

[44] P. Tsakalides and C. L. Nikias, *Deviation from normality in statistical signal processing: Parameter estimation with alpha-stable distributions*, A Practical Guide to Heavy Tails: Statistical Techniques and Applications, (1998), pp. 379–404.

[45] G. A. Tsihrintzis, *Statistical modeling and receiver design for multi-user communication networks*, A Practical Guide to Heavy Tails: Statistical Techniques and Applications, (1998).

[46] T. Wan, N. Canagarajah, and A. Achim, *Segmentation of noisy colour images using Cauchy distribution in the complex wavelet domain*, IET Image Processing, 5 (2011), pp. 159–170.

[47] F. Wang, W. Cao, and Z. Xu, *Convergence of multi-block Bregman ADMM for nonconvex composite problems*, arXiv preprint arXiv:1505.03063, (2015).

[48] Y. Wang, W. Yin, and J. Zeng, *Global convergence of ADMM in nonconvex nonsmooth optimization*, arXiv preprint arXiv:1511.06324, (2015).

[49] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, *Image quality assessment: from error visibility to structural similarity*, IEEE Transactions on Image Processing, 13 (2004), pp. 600–612.

[50] Y. Xu, T.-Z. Huang, J. Liu, and X.-G. Lv, *Split Bregman iteration algorithm for image deblurring using fourth-order total bounded variation regularization model*, Journal of Applied Mathematics, 2013 (2013).

[51] Y. Xu and W. Yin, *A block coordinate descent method for multi-convex optimization with applications to nonnegative tensor factorization and completion*, tech. rep., 2012.

[52] J. Yang, Y. Zhang, and W. Yin, *A fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data*, IEEE Journal of Selected Topics in Signal Processing, 4 (2010), pp. 288–297.

[53] L. Yang, T. K. Pong, and X. Chen, *Alternating direction method of multipliers for nonconvex background/foreground extraction*, arXiv preprint arXiv:1506.07029, (2015).

[54] X.-L. Zhao, F. Wang, and M. K. Ng, *A new convex optimization model for multiplicative noise and blur removal*, SIAM Journal on Imaging Sciences, 7 (2014), pp. 456–475.